Genotyping Microarray for the Detection of More Than 200 CFTR Mutations in Ethnically Diverse Populations

Iris Schrijver,* Eneli Oitmaa,[†] Andres Metspalu,^{†‡} and Phyllis Gardner[§]

From the Department of Pathology,* Stanford University School of Medicine, Stanford, California; Asper Biotech, Ltd.,[†] Tartu, Estonia; the Institute of Molecular and Cell Biology,[‡] University of Tartu/Estonian Biocentre, Tartu, Estonia; and the Department of Medicine,[§] Stanford University School of Medicine, Stanford, California

Cystic fibrosis (CF), which is due to mutations in the cystic fibrosis transmembrane conductance regulator gene, is a common life-shortening disease. Although CF occurs with the highest incidence in Caucasians, it also occurs in other ethnicities with variable frequency. Recent national guidelines suggest that all couples contemplating pregnancy should be informed of molecular screening for CF carrier status for purposes of genetic counseling. Commercially available CF carrier screening panels offer a limited panel of mutations, however, making them insufficiently sensitive for certain groups within an ethnically diverse population. This discrepancy is even more pronounced when such carrier screening panels are used for diagnostic purposes. By means of arrayed primer extension technology, we have designed a genotyping microarray with 204 probe sites for CF transmembrane conductance regulator gene mutation detection. The arrayed primer extension array, based on a platform technology for disease detection with multiple applications, is a robust, cost-effective, and easily modifiable assay suitable for CF carrier screening and disease detection. (*J Mol Diagn* 2005, 7:375–387)

Cystic fibrosis (CF) is a severe, common autosomal recessive disease due to mutations in the cystic fibrosis transmembrane conductance regulator (CFTR) gene (OMIM number *602421; <http://www.ncbi.nlm.nih.gov/Omim/>). Asymptomatic carrier parents, who have no physiological or biochemical outcome that enables routine identification, typically have one CFTR mutation; whereas diseased progeny carry at least two mutations, one on each CFTR gene allele. CF has a high incidence in people of Northern European descent, occurring in approximately 1 in 2500 live births.

Because of this, the American College of Medical Genetics and the American College of Obstetricians and Gynecologists (ACMG and ACOG) recently called for CF mutation carrier screening for expecting couples and for those contemplating pregnancy in high-risk groups (Northern European and Ashkenazi Jewish populations) as well as for proper counseling about diagnostic availability and limitations of screening in other groups with lower disease risk.¹ By use of the inclusion criterion of mutations having a threshold of 0.1% frequency in the general U.S. population, a routine screening panel of 25 mutations was initially selected from the more than 1300 currently described CFTR sequence variants (<http://www.genet.sickkids.on.ca/cftr/>). This recommended panel was recently modified and currently includes 23 mutations.² The selected panel chiefly reflects the mutations prevalent in the high-risk populations of Northern European and Ashkenazi Jewish descent. Mutations that are prevalent in other races and ethnicities are mostly not included in the routine CF 25 mutation panel because they do not reach inclusion criteria for the general population threshold.

Although development of commercial versions of the recommended screening panel (usually augmented by one to six mutations) has led to on-site CFTR mutation carrier screening, there are serious limitations of the currently available tests for widespread use, chiefly the detection rate, which depends on ethnicity. The risk of CF mutation carrier status varies by ethnicity, with the highest risk in people of Northern European ancestry at 1 in 25 and in people of Ashkenazi Jewish descent at 1 in 29.^{1,2} Other populations have discernible risk, however, including Hispanic Americans (1 in 46), African Americans (1 in 65), and Asian Americans (~1 in 90). The latter three estimates may be lower than the actual risk in these ethnicities because of ascertainment bias due to the misperception that CF is a "Caucasian disease." In addition, some ethnic groups (eg,

Supported by Stanford University Fund 1HLP 111 (to P.G.), and by ESF grant 4578 and Estonian Ministry of Research and Education grant 0182582s03 (to A.M.).

Accepted for publication February 28, 2005.

Address reprint requests to Iris Schrijver, M.D., Department of Pathology, L235, Stanford University Medical Center, 300 Pasteur Drive, Stanford, CA 94305. E-mail: iris.schrijver@medcenter.stanford.edu.

Asians) have not, as yet, been studied at the molecular level as thoroughly as the Caucasian populations. Thus, whereas the current on-site tests can achieve a detection rate of 90% in Northern European Caucasians with CFTR carrier status, the ability to detect CF mutation carrier status in other populations falls off considerably, with the estimated detection rates of 69% in African Americans and 57% in Hispanic CF carriers.¹

Currently, if the on-site carrier screening with the recommended panel of CFTR mutations is negative and if there is a family history or anxiety is high, DNA samples may be sent to reference laboratories for more comprehensive analysis that involves significantly higher costs. For non-Caucasians, carrier screening by a larger commercial panel is often preferentially chosen by counselors over the ACMG/ACOG recommended panel because of detection limitations. For diagnostic purposes in suspected CF patients, a carrier screening panel is sometimes used; but in those instances, second tier testing is frequently necessary. This second tier of more comprehensive testing generally involves either larger mutation panels or scanning methodologies, such as differential gradient gel electrophoresis, denaturing high-performance liquid chromatography, temporal temperature gradient gel electrophoresis, or single-strand conformation analysis, followed by direct DNA sequencing to characterize the mutations identified by scanning techniques.³ The gold standard of direct DNA gene sequencing is currently too costly for routine diagnostic use.

In light of the ethnic diversity of the U.S. population and the frequency of ethnically mixed marriages (<http://www.census.gov/>), we believe that a significantly expanded CF mutation panel, suitable for on-site use, would be beneficial as an option for routine CF mutation carrier screening. This may especially be desirable in non-Caucasians and individuals of mixed ethnicity. We describe a molecular diagnostic assay, using a new use of the arrayed primer extension (APEX) technology⁴ able to detect 204 different CFTR mutations that include a large number of mutations detected in non-Caucasians as well as those that are frequent in the Caucasian population and in the U.S. population as a whole. The assay is based on single-primer nucleotide extension, first described by Shumaker et al⁵ in 1996 and subsequently converted to an array format.⁴ It is a methodology that enables accurate mutation detection in a microarray format. Our CF APEX microarray described herein is suitable for carrier screening as well as molecular diagnosis of affected individuals.

Materials and Methods

Mutation Selection

The 204 mutations on the APEX microarray were selected from the CF Genetic Analysis Consortium (1994) (<http://www.genet.sickkids.on.ca/>) and the literature,⁶⁻⁹

representing the most frequently screened mutations in Caucasians and those identified as recurring in specific Caucasian and non-Caucasian populations. The full set of mutations is listed in Table 1. The sequence numbering in this manuscript is according to the CFTR GenBank reference sequence NM_000492 (<http://www.ncbi.nlm.nih.gov/GenBank/>).⁹

Oligonucleotide Microchips

Oligonucleotide primers were designed according to the wild-type CFTR gene sequence for both the sense and antisense directions. The 25-bp oligonucleotides with 6-carbon amino linkers at their 5' end were obtained from MWG (Munich, Germany). Most scanning oligonucleotides were designed to scan 1 bp in the wild-type sequence, except in the case of deletions and insertions that have the same nucleotide in the 1-bp direction. In this case, we designed the oligo to extend further into the deletion or insertion to enable discrimination of the nucleotide change. For example: 5'-AGCCTGGCACCATTAAAGAAAATATCAT-3' Δ F508 S; 5'-TTTCCTGGATTATGCCTGGCACCATTAAAGAAAATATCATCTTTGGTGTTTCCTATGATGAATATAGATACAGA-3'; Δ F508 AS 3'-AACACAAAGGATACTACTTATATC-5' in which A represents a deliberate mismatch to avoid strong secondary structure and CTT represents a deletion of three nucleotides. In case of the normal allele, we expect signals for the sense oligo in the cytosine (C) channel and for the antisense oligo in the adenine (A) channel (C/A). In case of the mutant allele, we will detect signals for the sense oligo in the thymine (T) channel. The signal corresponding to the antisense oligo will also appear in the T channel (T/T).

The microarray slides used for spotting the oligonucleotides have a dimension of 24 × 60 mm and are coated with 3-aminopropyl-trimethoxysilane plus 1,4-phenylenedi-isothiocyanate (Asper Biotech, Ltd., Tartu, Estonia). Primers were diluted to 50 μ mol/L in 100 mmol/L carbonate buffer, pH 9.0, and spotted onto the activated surface with BioRad VersArray (BioRad Laboratories, Hercules, CA). The slides were blocked with 1% ammonia solution and stored at 4°C until needed. Washing steps with 95°C distilled water (TKA, Toshvin Analytical, Germany) and 100 mmol/L NaOH were performed before the APEX reactions to reduce the background fluorescence and to avoid rehybridization of unbound oligonucleotides to the APEX slide.

Each selected CFTR sequence variant is identified by two unique 25-mer oligonucleotides, one for the sense and one for the antisense strand, although for some mutations, three oligonucleotides are used. On the other hand, fewer than expected oligonucleotides are used in some cases. For example, when different nucleotide substitutions (with a maximum of three) occur at the same nucleotide position, the accurate sequence can be determined using the same pair of detection oligos (sense and antisense). The APEX primers for mutations 3121-2A>G and 3121-2A>T, as an example, are the same, where A represents the location of the mutation under

Table 1. Complete List of Mutations Detectable with the CF APEX Assay

	CFTR location	Amino acid change	Nucleotide change
1	E 1	Frameshift	175delC
2	E 2,3	Frameshift	del E2, E3
3	E 2	W19C	189 G>T
4	E 2	Q39X	247 C>T
5	IVS 2	Possible splicing defect	296 + 12 T>C
6	E 3	Frameshift	359insT
7	E 3	Frameshift	394delTT
8	E 3	W57X (TAG)	302G>A
9	E 3	W57X (TGA)	303G>A
10	E 3	E60X	310G>T
11	E 3	P67L	332C>T
12	E 3	R74Q	353G>A
13	E 3	R75X	355C>T
14	E 3	G85E	386G>A
15	E 3	G91R	403G>A
16	IVS 3	Splicing defect	405 + 1G>A
17	IVS 3	Possible splicing defect	405 + 3A>C
18	IVS 3	Splicing defect	406 - 1G>A
19	E 4	E92X	406G>T
20	E 4	E92K	406G>A
21	E 4	Q98R	425A>G
22	E 4	Q98P	425A>C
23	E 4	Frameshift	444delA
24	E 4	Frameshift	457TAT>G
25	E 4	R117C	481C>T
26	E 4	R117H	482G>A
27	E 4	R117P	482G>C
28	E 4	R117L	482G>T
29	E 4	Y122X	498T>A
30	E 4	Frameshift	574delA
31	E 4	I148T	575T>C
32	E 4	Splicing defect	621G>A
33	IVS 4	Splicing defect	621 + 1G>T
34	IVS 4	Splicing defect	621 + 3A>G
35	E 5	Frameshift	624delT
36	E 5	Frameshift	663delT
37	E 5	G178R	664G>A
38	E 5	Q179K	667C>A
39	IVS 5	Splicing defect	711 + 1G>T
40	IVS 5	Splicing defect	711 + 1G>A
41	IVS 5	Splicing defect	712 - 1G>T
42	E 6a	H199Y	727C>T
43	E 6a	P205S	745C>T
44	E 6a	L206W	749T>G
45	E 6a	Q220X	790C>T
46	E 6b	Frameshift	935delA
47	E 6b	Frameshift	936delTA
48	E 6b	N287Y	991A>T
49	IVS 6b	Splicing defect	1002 - 3T>G
50	E 7	ΔF311	3-bp del between nucleotides 1059 and 1069
51	E 7	Frameshift	1078delT
52	E 7	Frameshift	1119delA
53	E 7	G330X	1120G>T
54	E 7	R334W	1132C>T
55	E 7	I336K	1139T>A
56	E 7	T338I	1145C>T
57	E 7	Frameshift	1154insTC
58	E 7	Frameshift	1161delC
59	E 7	L346P	1169T>C
60	E 7	R347H	1172G>A
61	E 7	R347P	1172G>C
62	E 7	R347L	1172G>T
63	E 7	R352Q	1187G>A
64	E 7	Q359K/T360K	1207C>A and 1211C>A
65	E 7	S364P	1222T>C
66	E 8	Frameshift	1259insA
67	E 8	W401X (TAG)	1334G>A
68	E 8	W401X (TGA)	1335G>A
69	IVS 8	Splicing changes	1342 - 6 poly(T) variants 5T/7T/9T
70	IVS 8	Splicing defect	1342 - 2A>C

Table 1. Continued

	CFTR location	Amino acid change	Nucleotide change
71	E 9	A455E	1496C>A
72	E 9	Frameshift	1504delG
73	E 10	G480C	1570G>T
74	E 10	Q493X	1609C>T
75	E 10	Frameshift	1609delCA
76	E 10	ΔI507	3-bp del between nucleotides 1648 and 1653
77	E 10	ΔF508	3-bp del between nucleotides 1652 and 1655
78	E 10	Frameshift	1677delTA
79	E 10	V520F	1690G>T
80	E 10	C524X	1704C>A
81	IVS 10	Possible splicing defect	1717 – 8G>A
82	IVS 10	Splicing defect	1717 – 1G>A
83	E 11	G542X	1756G>T
84	E 11	G551D	1784G>A
85	E 11	Frameshift	1784delG
86	E 11	S549R (A>C)	1777A>C
87	E 11	S549I	1778G>T
88	E 11	S549N	1778G>A
89	E 11	S549R (T>G)	1779T>G
90	E 11	Q552X	1786C>T
91	E 11	R553X	1789C>T
92	E 11	R553G	1789C>G
93	E 11	R553Q	1790G>A
94	E 11	L558S	1805T>C
95	E 11	A559T	1807G>A
96	E 11	R560T	1811G>C
97	E 11	R560K	1811G>A
98	IVS 11	Splicing defect	1811 + 1.6 kb A>G
99	IVS 11	Splicing defect	1812 – 1G>A
100	E 12	Y563D	1819T>G
101	E 12	Y563N	1819T>A
102	E 12	Frameshift	1833delT
103	E 12	D572N	1846G>A
104	E 12	P574H	1853C>A
105	E 12	T582R	1877C>G
106	E 12	E585X	1885G>T
107	IVS 12	Splicing defect	1898 + 5G>T
108	IVS 12	Splicing defect	1898 + 1G>A
109	IVS 12	Splicing defect	1898 + 1G>C
110	IVS 12	Splicing defect	1898 + 1G>T
111	E 13	Frameshift	1924del7
112	E 13	del of 28 amino acids	1949del84
113	E 13	I618T	1985T>C
114	E 13	Frameshift	2183AA>G
115	E 13	Frameshift	2043delG
116	E 13	Frameshift	2055del9>A
117	E 13	D648V	2075T>A
118	E 13	Frameshift	2105–2117 del13insAGAA
119	E 13	Frameshift	2108delA
120	E 13	R668C	2134C>T
121	E 13	Frameshift	2143delT
122	E 13	Frameshift	2176insC
123	E 13	Frameshift	2184delA
124	E 13	Frameshift	2184insA
125	E 13	Q685X	2185C>T
126	E 13	R709X	2257C>T
127	E 13	K710X	2260A>T
128	E 13	Frameshift	2307insA
129	E 13	V754M	2392G>A
130	E 13	R764X	2422C>T
131	E 14a	W846X	2670G>A
132	E 14a	Frameshift	2734delGinsAT
133	E 14b	Frameshift	2766del8
134	IVS 14b	Splicing defect	2789 + 5G>A
135	IVS 14b	Splicing defect	2790 – 2A>G
136	E 15	Q890X	2800C>T
137	E 15	Frameshift	2869insG
138	E 15	S945L	2966C>T
139	E 15	Frameshift	2991del32
140	E 16	Splicing defect	3120G>A

Table 1. Continued

	CFTR location	Amino acid change	Nucleotide change
141	IVS 16	Splicing defect	3120 + 1G>A
142	IVS 16	Splicing defect	3121 - 2A>G
143	IVS 16	Splicing defect	3121 - 2A>T
144	E 17a	Frameshift	3132delTG
145	E 17a	I1005R	3146T>G
146	E 17a	Frameshift	3171delC
147	E 17a	Frameshift	3171insC
148	E 17a	del V1022 and I1023	3199del6
149	E 17a	Splicing defect	3271delGG
150	IVS 17a	Possible splicing defect	3272 - 26A>G
151	E 17b	G1061R	3313G>C
152	E 17b	R1066C	3328C>T
153	E 17b	R1066S	3328C>A
154	E 17b	R1066H	3329G>A
155	E 17b	R1066L	3329G>T
156	E 17b	G1069R	3337G>A
157	E 17b	R1070Q	3341G>A
158	E 17b	R1070P	3341G>C
159	E 17b	L1077P	3362T>C
160	E 17b	W1089X	3398G>A
161	E 17b	Y1092X (TAA)	3408C>A
162	E 17b	Y1092X (TAG)	3408C>G
163	E 17b	L1093P	3410T>C
164	E 17b	W1098R	3424T>C
165	E 17b	Q1100P	3431A>C
166	E 17b	M1101K	3434T>A
167	E 17b	M1101R	3434T>G
168	IVS 17b	3500 - 2A>T	3500 - 2A>T
169	IVS 17b	Splicing defect	3500 - 2A>G
170	E 18	D1152H	3586G>C
171	E 19	R1158X	3604C>T
172	E 19	R1162X	3616C>T
173	E 19	Frameshift	3659delC
174	E 19	S1196X	3719C>G
175	E 19	S1196T	3719T>C
176	E 19	Frameshift and K1200E	3732delA and 3730A>G
177	E 19	Frameshift	3791delC
178	E 19	Frameshift	3821delT
179	E 19	S1235R	3837T>G
180	E 19	Q1238X	3844C>T
181	IVS 19	Possible splicing defect	3849 + 4A>G
182	IVS 19	Splicing defect	3849 + 10 kb C>T
183	IVS 19	Splicing defect	3850 - 1G>A
184	E 20	G1244E	3863G>A
185	E 20	G1244V	3863G>T
186	E 20	Frameshift	3876delA
187	E 20	G1249E	3878G>A
188	E 20	S1251N	3884G>A
189	E 20	T1252P	3886A>C
190	E 20	S1255X	3896C>A and 3739A>G in E19
191	E 20	S1255L	3896C>T
192	E 20	Frameshift	3905insT
193	E 20	D1270N	3940G>A
194	E 20	W1282R	3976T>C
195	E 20	W1282X	3978G>A
196	E 20	W1282C	3978G>T
197	E 20	R1283M	3980G>T
198	E 20	R1283K	3980G>A
199	IVS 20	Splicing defect	4005 + 1G>A
200	E 21	Frameshift	4010del4
201	E 21	Frameshift	4016insT
202	E 22	Inframe del E21	del E21
203	E 21	N1303K	4041C>G
204	E 24	Frameshift	4382delA

interrogation: ACCAACATGTTTTCTTTGATCTTAC 3121-2A→G,T S; 5'-ACCAACATGTTTTCTTTGATCTTAC A GTTGTTATTAATTGTGATTGGAGCTATAG-3'; CAACAA-

TAATTAACACTAACCTCGA 3121-2A→G,T AS. These variables result in a total of 379 oligonucleotides annealed to the microarray slide.

Table 2. Genomic DNA Samples Used for Mutation Evaluation on the APEX Array

Mutations validated with native DNA
CFTRdel 2,3 (21 kb)
394delTT
G85E
R75X
574delA
Y122X
R117C
R117H
621 + 1G>T
621 + 3A>G
711 + 1G>T
I336K
R334W
R347P
IVS8-5T
IVS8-7T
IVS8-9T
A455E
ΔF508
ΔI507
1677delTA
1717 - 1G>A
G542X
G551D
R553X
R560T
S549N
1898 + 1G>A
1898 + 1G>C
2183AA>G
2043delG
R668C
2143delT
2184delA
2184insA
2789 + 5G>A
S945L
3120 + 1G>A
I1005R
3272 - 26A>G
R1066C
G1069R
Y1092X (C>A)
3500 - 2A>T
R1158X
R1162X
3659delC
S1235R
3849 + 10 kb C>T
W1282X

Genomic and Synthetic Template Samples

Where possible, native genomic DNA was collected. Thus, 51 patient or cell line samples with known mutations were evaluated on the chip (Table 2). The presence of mutations in commercially available samples (Coriell Cell repositories; <http://locus.umdj.edu/ccr/>) was verified in the Molecular Pathology laboratory at Stanford Hospitals and Clinics. Additional samples were obtained from the Molecular Diagnostics Centre of United Laboratories at Tartu University Clinics. Some DNA samples were a generous gift from Dr. Milan Macek, Jr., at Charles University. This set of samples was anonymized. When it was not readily possible to obtain native genomic DNA

samples containing the screened mutations, synthetic approximately 50-bp templates were designed according to the mutated CFTR sequence for both the sense and antisense directions and optimized for melting temperature (MWG). In this case, poly(T) tracts were designed at the 5' end to minimize the possibility of the self-extensions and/or self-annealing of the synthetic templates.

Template Preparation

The CFTR gene was amplified from genomic DNA in 30 amplicons. The PCR reaction mixture (50 μ l) was optimized with the following: 10 $\times$ TaqDNA polymerase buffer; 2.5 mmol/L MgCl₂ (Naxo, Estonia); 0.25 mmol/L dNTP (MBI Fermentas, Vilnius, Lithuania) (20% fraction of dTTP was substituted with dUTP); and 10 pmol of primer stock, genomic DNA (approximately 80 ng), SMART-Taq Hot DNA polymerase (3 U) (Naxo, Estonia), and sterile deionized water. After amplification (MJ Research DNA Thermal Cycler; MJ Research, Inc., Waltham, MA), the amplification products were concentrated and purified using Jetquick spin columns (Genomed GmbH, Lohne, Germany). In a one-step reaction, the functional inactivation of the traces of unincorporated dNTPs was achieved by addition of shrimp alkaline phosphatase (Amersham Pharmacia Biotech, Inc., Milwaukee, WI), and fragmentation of the PCR product was achieved by the addition of thermolabile uracil *N*-glycosylase (Epicenter Technologies, Madison, WI) followed by heat treatment.⁴

APEX Reactions

The APEX mixture consisted of 32 μ l of fragmented product, 5 U of Thermo Sequenase DNA polymerase (Amersham Pharmacia Biotech), 4 μ l of Thermo Sequenase reaction buffer (260 mmol/L Tris-HCl, pH 9.5, and 65 mmol/L MgCl₂) (Amersham Pharmacia Biotech) and 1 μ mol/L final concentration of each fluorescently labeled ddNTPs: Cy5-ddUTP, Cy3-ddCTP, Texas Red-ddATP, Fluorescein-ddGTP, (PerkinElmer Life Sciences, Wellesley, MA). The DNA was first denatured at 95°C for 10 minutes. The enzyme and the dyes were immediately added to the DNA mixture, and the whole mixture was applied to prewarmed slides (58°C). The reaction was allowed to proceed for 10 minutes at 58°C, followed by washing once with 0.3% Alconox (Alconox, Inc.) and twice for 90 seconds at 95°C with distilled water (TKA, Germany). A droplet of antibleaching reagent (AntiFade SlowFade; Molecular Probes Europe BV, Leiden, The Netherlands) was applied to the slides before imaging. This process is depicted schematically in Figure 1.

During assay development, a few of the oligonucleotides designed from the wild-type CFTR sequence failed to perform the APEX reaction. The chief reason for APEX primer failure is the formation of self-annealing secondary structures that fail to hybridize or facilitate self-priming and extension. To obviate this problem, we designed new versions of such primers by incorporating a mismatch or a modified nucleotide at the 5' or internal part of the

Figure 1. Schematic depiction of the APEX reaction.

primer. Such changes can reduce primer self-complementarity without compromising hybridization and primer extension. In the case of secondary structures at the 3' end, which is essential for template annealing and extension, some alternative versions with internal base substitutions can be attempted, but not all work. After our final design of the APEX primers, 184 sequence variants (out of 204 variants at 181 sites) were detected in both of the sense and antisense directions, 7 from only the sense strand (the antisense strand does not work reliably), and 13 from only the antisense strand (the sense strand does not work reliably).

Analysis

The APEX array had four points of data analysis: both the forward and reverse primers are spotted in duplicate. This vastly reduced the possibility of false-positive signal interpretation due to dust particles or other nonspecific reactions and allowed distinction between homo- and heterozygosity. The array images were captured by means of detector Genorama QuattroImager 003 (Asper Biotech) at 20- μ m resolution. The device combines a total internal reflection fluorescence-based excitation mechanism with a charge-coupled device camera.⁴ Sequence variants were identified using Genorama 3.0 genotyping software.

Results

We selected 204 CFTR sequence variants for a comprehensive diagnostic panel to increase CF carrier and disease detection across all racial and ethnic groups (Table 1). Sequence variants selected were obtained from the CF Genetic Analysis Consortium (1994) that compiled information from the screening of 43,849 chromosomes and from studies of relatively small-size samples or those isolated to specific ethnic populations (<http://www.genet.sickkids.on.ca/cgi-bin/WebObjects/MUTATION>). The population frequencies of the included mutations vary considerably according to ethnicity and number of

Table 3. PCR Primers Used for Amplification of the Coding Sequence and Splice Sites in the CFTR Gene

Primer name	Sequence 5' → 3'	Reference no.
1 F	TCTTTGGCATTAGGAGCTTG	
1 R	CAAACCCCAACCCATACACAC	
IVS 1 F	GTTGTAAATCCTGGACCTGCAGAAG	
IVS 1 R	CCCTCCTCTGATTCCACAAGGTAT	
2 F	TCCATATGCCAGAAAAGTTG	
2 R	ACAAGCATGCACCTACCATTG	
3 F	CTTGGGTAAATCTCCTTGGAT	
3 R	CACCTATTCAACAGATTTTCG	
IVS 3 F	CCTTAGCAATTTGTATGAGCCAA	
IVS 3 R	CCATCATAGGATACAATGAATGCTGG	
4 F	CCACTGTTGCTATAACAAATCCCAA	
4 R	TTCAGCATTATCCCTTACTTGTACC	
5 F	ATTTCTGCTAGATGCTGGG	9
5 R	AACTCCGCCTTTCCAGTTGT	9
6a F	GCTCAGAACCCAGAAAGTTT	
6a R	CATCATCATCTCCCTAGCC	
6b F	GGCACAATAGGAGGCATTTACCAA	
6b R	GTCACAAACATCAAAATGAGGTGGA	
7 F	GACCATGCTCAGATCTTCCATTCC	
7 R	CCACTCTCATCCATCATACTGTCCA	
8 F	AGTGGGTAAATCAGGGTTGCTTTG	
8 R	CACCTGGCCATTCTCTACTTCTTT	
9 F	GGCCATGTGCTTTTCAATC	
9 R	CTCCAAAAATACCTTCCAGCAC	
10 F	TGTGCCCTTCTCTGTGAACCTCTA	
10 R	TGATCCATTACAGTAGCTTACCCA	
11 F	TGTGGTTAAAGCAATAGTGTGA	
11 R	ACAGCAAATGCTTGTCTAGAC	
IVS 11 F	TGTAGCTCTTGCATCTGTCTTC	
IVS 11 R	AGTGCTGCCACAACCTGTATAA	
12 F	CTTCAGCAGTTCTTGGAAATGTTG	
12 R	GGAAGTAATCTTGAATCCTGGCCC	
13 F	ATGCTAAATACGAGACATATTGC	
13 R	ACATGCTACATATTGCATTCTACTC	
14a F	CATTATTCTGGCTATAGAATGACA	
14a R	TGTATACATCCCCAAATATCTTA	
14b F	GCATGGGAGGAATAGGTGAA	
14b R	TGCTTGGGAGAAATGAAACA	
15 F	GTGCAGCTCTTGCAGTTTCTAAAG	
15 R	ACCAACAAAACACAGGCCCTA	
16 F	GCAGGATGAGTACCCACCTATTCTT	
16 R	GAAAGATGGTCTTTGTGCTCTCT	
17a F	CACTGACACACTTTGTCCACTTTGC	
17a R	GAATGTCCTGTACACCAACTGTGG	
17b F	TATTCAAAGAATGGCACCAGTGTG	
17b R	GACAATCTGTGTGATCGGTTTTT	
18 F	GTGCCCTAGGAGAAGTGTGA	
18 R	GACAGATACACAGTGACCCCTCA	
19 F	CCTGTTAGTTCATTGAAAAGCCCCG	
19 R	TTCTGCTAACACATGTCTCAGGC	
IVS 19 F	GATTCTGGAGACCACAAGG	
IVS 19 R	ACCTCCTCCCTGAGAATGTT	
20 F	ATCTTCCACTGGTGACAGGA	
20 R	CTGCCTATGAGAAAACCTGCAC	
21 F	AATGTTCAACAGGGACTCCA	9
21 R	CCTGTTGCTCCAGGTATGTT	
24 F	TGCCTTCTGTCCCAGATCTCACTA	
24 R	TCAGTGTCTCAATTCCCTTACC	

samples screened, but they represent the most common reported mutations across population groups to date. Mutations include several prominent in non-Caucasian backgrounds, including N1303K, 3849 + 10 kb, 2789 + 5G>A, 3876delA, 406-1G>A, R75X, 2055delG>A, and S549N, which are each prevalent in the Hispanic popu-

Figure 2. APEX analysis at mutation site 2183AA>G. Each numbered **row** represents the analysis of an individual patient sample. The **row** presents two sets of four-channel fluorescent images representing the bases adenine (A), cytosine (C), guanine (G), and thymine (T), respectively for the sense strand (**top**) and antisense strand (**bottom**). The histograms to the right of the fluorescent images are of the fluorescent intensities of the four channels at the mutation analysis site. The letters to the right of the histogram represent the base(s) identified on each strand. **Row 3** presents the results of heterozygous target DNA derived from a CF patient (WT/2183AA>G). In this case, the sense strand is extended by both the wild-type (WT) complementary target sequence base A and the base G complementary for the mutation, whereas the antisense strand is extended by the WT base T and the base C complementary for the mutation. **Row 4** contains the results of normal DNA derived from a non-CF individual at the target sequence (WT/WT), with the expected WT base A in the sense channel and WT base T in the antisense channel. **Row 5** contains the results of a homozygous target DNA derived from a CF patient (2183AA>G/2183AA>G), with the base G complementary for the mutation in the sense channel and the base C complementary for the mutation in the antisense channel.

lation (personal observation),^{7,10-13} 3120 + 1G>A, which is prevalent in the African-American population; and 1898 + 5G>T, which is prevalent in the Chinese population (<http://www.genet.sickkids.on.ca/cftr/>). The mutations originate from 26 exons (exons 1 to 22, including exons 6a and 6b, 14a and 14b, and 17a and 17b, and exon 24) and from 15 introns (2, 3, 4, 5, 6b, 8, 10, 11, 12, 14b, 16, 17a, 17b, 19, and 20). They include single-nucleotide substitutions, technically the most easy to detect with the APEX reaction, as well as insertions, deletions, including the large 21-kb deletion, which removes

exons 2 and 3, and even repeats, such as the 5T/7T/9T repeats important in congenital bilateral absence of the vas deferens (CBAVD). Sample DNA is amplified in 30 amplicons with 29 pairs of PCR primers (Table 3) encompassing the mutations, with PCR mixtures that include 20% substitution of dUTPs for dTTPs, allowing for later fragmentation with uracil *N*-glycosylase as described previously.⁴

The general applicability of a proposed diagnostic will be determined by its clinical relevance, ease of use, low cost, and number of samples that can be processed per

Figure 3. APEX analysis at mutation site ΔF508. Three patient samples, one each with ΔF508/ΔF508, WT/ΔF508, and WT/WT, are presented. The results are presented as described in Figure 1. **Row 3 (top)** contains the results of homozygous target DNA derived from a CF patient (ΔF508/ΔF508). In this case, both the sense and the antisense strands are extended by the base T complementary in both strands for the mutation. **Row 10 (center)** contains the results of heterozygous target DNA from a CF patient (WT/ΔF508). In this case, the sense strand is extended by both the WT base C and base T complementary for the mutation, whereas the antisense strand is extended by the WT base A and the base T complementary for the mutation. **Row 11** contains the results from a non-CF individual (WT/WT), in which the sense strand is extended by the WT base C, and the antisense strand is extended by the WT base A.

run. In general, the entire APEX reaction process, from PCR amplification (2 hours), PCR product purification (20 minutes), DNA fragmentation (1 hour), APEX reaction (15 minutes), to visualization of results (6 minutes) and analysis of results (10–15 minutes) can be accomplished in 4 to 5 hours. The time for sample preparation will be minimized in the future with the development of automated sample preparation and analysis. Because all steps leading up to the visualization of results on the Quattrolmager can be performed in parallel, the rate-limiting factor in sample analysis will be determined by the number of imagers per test site. At 6 to 7 minutes analysis time per array, the throughput in 8 hours would be approximately 60 to 70 patients per imager unit.

The results allow reliable and reproducible detection of the wild-type (WT) versus mutation sequence at each array position on the APEX CF microarray. The assay is suitable both for screening of CF carriers (one heterozygous mutation in entire array) and for diagnosis of patients (two mutations, either heterozygous at two array sites or homozygous at one array site). Representative results are presented in Figures 2 and 3. Figure 2 demonstrates results from three patient samples, one each of normal, heterozygous, and homozygous, at the grid position for mutation 2183AA>G, located in exon 13 (R domain). This mutation has a frequency of 2% in a screened sample of 2006 European probands (<http://www.genet.sickkids.on.ca/cgi-bin/WebObjects/MUTATION>). Figure 3 shows the results from three patient samples at the grid position for the common mutation $\Delta F508$, again one each for normal, heterozygous, and homozygous at that position. In each case, the results accurately detect the sequence of both alleles for each patient sample. The 5T/7T/9T poly(T) tract polymorphism in IVS 8, of which the 5T variant is particularly important in CBAVD, is the most technically difficult mutation to accurately identify. Differentiation of the poly(T) length requires three sets of APEX primers for accurate detection. Representative results for three patient samples (5T/7T, 7T/9T, and 9T/9T) are presented in Figure 4. The first set of primers can distinguish between 5T and 7T/9T. The second set can distinguish between 5T and 9T. The third set of primers can distinguish between 7T and 9T. The first two of the three primer pairs present a technical difficulty previously alluded to, ie, despite several attempts to redesign and optimize the allele-specific primer pairs, an APEX oligo does not necessarily function reliably in both directions. Specifically the sense strand for the first primer pair and the antisense strand for the second primer pair are unreliable, and results from these must be disregarded. The antisense strand for the first pair and the sense strand for the second pair provide reliable identification of the probed base sequence, however, in each case. By testing at all three allele-specific oligo pairs for the IVS8 5T/7T/9T site, the length of the T repeat can be accurately and reproducibly determined.

The CF APEX microarray was validated in a pilot study by means of 51 patient samples (Table 2) with different CFTR mutations. Mutation sites in which relevant patient samples could not readily be obtained were tested by

Figure 4. APEX analysis at mutation site IVS8-5T/7T/9T. Representative results for three patient samples are shown at mutation site IVS8-5T/7T/9T. This challenging mutation site requires three pairs of allele-specific primers for accurate identification. The first set of primers (**A**) consists of a sense strand that does not work reliably despite several iterations and thus should be discounted and an antisense strand predicted to give base C for 5T (–/C) and base A for either 7T or 9T (–/A). The second set of primers (**B**) consists of a sense strand that elongates with a C only for 9T and an antisense strand that extends only with a C only for 5T. Thus the expected results for this set of primers is 5T (–/C), 7T (–/–), and 9T (C/–). The third set of primers consists of a sense strand oligo that extends with A for 7T and T for 9T, whereas the antisense strand extends with C for 7T and A for 9T. Thus the expected set of results for the third set of primers is 5T (–/–), 7T (A/C), and 9T (T/A). Adding the three sets of results together, patient sample 30 can be identified as compound heterozygous 5T/7T, patient sample 31 as compound heterozygous 7T/9T, and patient sample 32 as homozygous 9T/9T.

means of 136 synthetic primers (Table 4), designed as approximately 50-mer oligonucleotides based on the wild-type sequence but incorporating the mutation to be identified. Four sites were tested with patient samples and synthetic template DNA with highly comparable results. With this pilot validation series, sensitivity (TP/TP+FN) was 100%. No false negatives were observed. Specificity (TN/TN + FP) was 100% and no false positives were present. The APEX reactions are entirely reproducible under our testing conditions. Testing at each site occurred from 3 to 20 times with identical results, as long as the following requirements have been met: 1) good quality of the DNA sample; 2) clean PCR products, identified as a single band of expected size on agarose gels; 3) complete fragmentation of the uracil *N*-glycosylase-treated PCR product; and 4) optimization and validation of the arrayed oligos. The reported results constitute a pilot demonstration of the effectiveness of the APEX approach in CF mutation detection. Complete validation of the CF APEX chip, including true sensitivity and specificity, will require systematic evaluation of patient samples with mutations at each of the array sites, and studies performed in a blinded manner.

Table 4. Primers Generated to Create Synthetic Templates That Serve As Positive Mutation Controls

Primer name	Sense strand 5' → 3'	Name	Antisense strand 5' → 3'
175delC synt F	T (15) ATTTT TTTCAGGTCAGAGGTTGGCCA	175delC synt R	T (15) ATTT TGGAGACAACGCTGGCCCTTTTCC
W19C synt F	T (15) TACC AGACCAATTTTGAGGAAAGGAT	W19C synt R	T (15) ACAG CTAAAAATAAGAGAGGAGGAAC
Q39X synt F	T (15) TAAAT CCCTTCTGTTGATTCTGCTGA	Q39X synt R	T (15) AGTAT ATGTCTGCAATATCCAGGCGC
296 + 12T>C synt F	T (15) CACAT TGTTTAGTTGAAGAGAGAAAT	296 + 12T>C synt R	T (15) GCAT GAACATACCTTTCCAATTTTTC
359insT synt F	T (15) TTTTT TTCTGGAGATTATGTTCTAT	359insT synt R	T (15) AAAAA ACATCGCCGAAGGGCATTAA
E60X synt F	T (15) TAGCT GGCTTCAAAGAAAAATCCTAA	E60X synt R	T (15) ATCT ATCCCCATCTCTGCAAAAGAAT
P67L synt F	T (15) TTAA ACTCATTATATGCCCTTCGGCGA	P67L synt R	T (15) AGAT TTTTCTTTGAAGCCAGCTCTCT
R74Q synt F	T (15) AGCG ATGTTTTTCTGGAGATTATG	R74Q synt R	T (15) TGA AGGCATTAATGAGTTTAGGATT
R75X synt F	T (15) TGAT TTTTTCTGGAGATTATGTT	R75X synt R	T (15) ACCG AAGGCATTAATGAGTTTAGGA
W57X(TAG) synt F	T (15) AGGA TAGAGAGCTGGCTTCAAAGAAA	W57X(TAG) synt R	T (15) TATT CTCTGCAAAAGAATAAAAGTG
W57X(TGA) synt F	T (15) AGAT AGAGAGCTGGCTTCAAAGAAA	W57X(TGA) synt R	T (15) TCAT TCTCTGCAAAAGAATAAAAGT
G91R synt F	T (15) AGGT TAAGGATCTCATTTGTACATTC	G91R synt R	T (15) TTA AATATAAAAAGATTCCATAGAAC
405 + 1G>A synt F	T (15) ATA AGGATCTCATTTGTACATTCAT	405 + 1G>A synt R	T (15) TCCC TAAATATAAAAAGATTCCATAG
405 + 3A>C synt F	T (15) CAG ATCTCATTTGTACATTCAT	405 + 3A>C synt R	T (15) GAC CCCTAAATATAAAAAGATTCCAT
406 – 1G>A synt F	T (15) AGA AGTCACCAAGCAGTACAGCCTC	406 – 1G>A synt R	T (15) TTA CAAAAGGGGAAAAACAGAGAAAT
E92X synt F	T (15) TAAG TCACCAAGCAGTACAGCCTCT	E92X synt R	T (15) ACT ACAAAAGGGGAAAAACAGAGAAA
E92K synt F	T (15) AAAG TCACCAAGCAGTACAGCCTCT	E92K synt R	T (15) TC TACAAAAGGGGAAAAACAGAGAAA
444delA synt F	T (15) GAT CATAGCTTCCATGACCCGGATA	444delA synt R	T (15) ATCT TCCCAGTAAGAGAGGCTGTACT
574delA synt F	T (15) CTT GGAATGCAGATGAGAATAGCTAT	574delA synt R	T (15) AGT GATGAAGGCCAAAAATGGCTGGG
621G>A synt F	T (15) AGT AATACTTCCCTTGACAGGCCCA	621G>A synt R	T (15) TTT CTTATAAATCAAATAAACATAG
Q98P synt F	T (15) CGC CTCTCTTACTGGGAAGATCATATA	Q98P synt R	T (15) GGT ACTGCTTTGGTGACTTCTCATAA
457TAT>G synt F	T (15) GGAC CCGGATAACAAGGAGGAACGCT	457TAT>G synt R	T (15) CGGA AGCTATGATTCTTCCCAGTAAG
I148T synt F	T (15) CTG GAATGCAGATGAGAATAGCTATG	I148T synt R	T (15) GTG TATGAAGGCCAAAAATGGCTGG
624delT synt F	T (15) CTT AAAGCTGTCAAGCCGTGTTCTAG	624delT synt R	T (15) TAAG TCTAAAAGAAAAATGGAAAGTT
663delT synt F	T (15) ATG GACAACTTGTTAGTCTCTCTTCC	663delT synt R	T (15) CAT ACTTATTTTATCTAGAACACGGC
G178R synt F	T (15) AGAC AACTTGTTAGTCTCTCTTCCAA	G178R synt R	T (15) TA ACTTATTTTATCTAGAACACGG
Q179K synt F	T (15) AAAC TGTTAGTCTCTCTTCCAA	Q179K synt R	T (15) TTCC AATACTTATTTTATCTAGAACAC
711 + 5G>A synt F	T (15) ATAC TATTTGATTTAATCTTTTAGGC	711 + 5G>A synt R	T (15) TTA TACTTCATCAAAATTTGTTCCAGT
712 – 1G>T synt F	T (15) TGG ACTTGCATTTGGCACATTCGTGT	712 – 1G>T synt R	T (15) TAT TGAAAATAAAGACACAGCAAAAC
H199Y synt F	T (15) TAT TTGCGTGGATCGCTCCTTTGCA	H199Y synt R	T (15) TAT GCCAATGCTAGTCCCTGAAAAATA
P205S synt F	T (15) TCT TTGCAAGTGGCAGTCTCATGGG	P205S synt R	T (15) TAA GCGATCCACAGCAATGTTGCCAAT
L206W synt F	T (15) GGC AAAGTGGCAGTCTCATGGGGCTA	L206W synt R	T (15) TCA AGGAGCGATCCACAGCAATGTC
Q220X synt F	T (15) TAG GCGTCTGCTTCTTGGGATTTGG	Q220X synt R	T (15) TAT TAAACATCCAGATTGAGCCCATG
936delTA synt F	T (15) AA TCCAATCTGTAAAGGCATACTGCT	936delTA synt R	T (15) TGA TTTTCAATCATTTCTGAGGTAATC
935delA synt F	T (15) GAA ATATCCAATCTGTAAAGCATATC	935delA synt R	T (15) TAT TTCAATCATTTCTGAGGTAATC
N287Y synt F	T (15) TACT TAAAGACAGTAAGTTGTTCCAA	N287Y synt R	T (15) TAT TCAATCATTTTTCATTTGCTTCT
1002 – 3T>G synt F	T (15) GAGA ACGAACTGAACTGACTCGGGA	1002 – 3T>G synt R	T (15) TCT AAAAACAATAACAATAAAATTTCA
1154insTC syntwt F	T (15) ATCT CATTCTGCATTGTTCTGCGCAT	1154insTC syntwt R	T (15) TTG AGATGGTGGTGAATATTTTCCGGA
1154insTC syntmt F	T (15) TCT CTCATCTGCATTGTTCTGCGCAT	1154insTC syntmt R	T (15) TAG AGATGGTGGTGAATATTTTCCGGA
DF311 mt syntV1 F	T (15) CTT CTTCTCAGGGTCTTTTGTGGTG	dF311 mt syntV1 R	T (15) GAGA GAAGGCTGAGCTATTGAAGTATC
G330X synt F	T (15) TGA ATCATCTCTCGGAAAAATTTAC	G330X synt R	T (15) ATT TGATTAGTGATAGGGAAGCACA
S364P synt F	T (15) CCT CTTGAGCAATAAACAATAACATA	S364P synt R	T (15) GGT CATACCATGTTTGTACAGCCAG
Q359K/T360K mt synt F	T (15) AAAA ATGGTATGACTCTCTTGGAGC	Q359K/T360K mt synt R	T (15) TTTT TTACAGCCAGGGAATTTGCCG
1078delT synt F	T (15) CTT GTGGTGTTTTATCTGTGCTTCC	1078delT synt R	T (15) CAAG AACCTTGAGAAGAAGAGGCTG
1119delA synt F	T (15) CAAG GAATCATCTCTCGGAAAAATTT	1119delA synt R	T (15) CTT GATTAGTGATAGGGAAGCAGC
1161delC synt F	T (15) GAT TGTTCTGCGCATGGCGCTGCTC	1161delC synt R	T (15) TC AGAATGAGATGGTGGTGAATTTT
T338I synt F	T (15) TCAC CATCTCATTTCTGCATTGTTCTG	T338I synt R	T (15) ATGA ATATTTTCCGAGGATGATTCC
R352Q synt F	T (15) AGCA ATTTCCCTGGGCTGTACAAAACA	R352Q synt R	T (15) TGA GTGACCCGCTATGCGAGAACAA
L346P synt F	T (15) CGCG CATGGCCGTCACCTCGGCAATTT	L346P synt R	T (15) GGA CAATGCAAGATGAGATGGTGGT
1259insA synt F	T (15) AAAA AGCAAGAAATATAAGACATTTGGA	1259insA synt R	T (15) TTTT TTGTAAGAAATCTTATTTATAAA
W401X(TAG)mtsnt F	T (15) AGGA GAGGTGAGAAATTTTAAAAAA	W401X(TAG)mtsnt R	T (15) TAGA AGGCTGTTACATTTCTCCATCAC
W401X(TGA) synt F	T (15) AGAG GAGGTGAGAAATTTTAAAAAA	W401X(TGA) synt R	T (15) TC AGAAGGCTGTTACATTTCTCCATCA
1342 – 2A>C synt F	T (15) CGGA TATTTGGGGAATTTATTTAGAAA	1342 – 2A>C synt R	T (15) GGT TAAAAAACAACACACACACAC
1504delG synt F	T (15) TGAT CCACTGTAGCAGGCAAGGTAGT	1504delG synt R	T (15) TCAG CAACCCGCAACCACTGTCTCT
G480C synt F	T (15) TGTA AAATTAAGCACAGTGAAGAAT	G480C synt R	T (15) ACT CTGAAGGCTCCAGTTCTCCCAT
C524X synt F	T (15) ACA ACTAGAGAGGTAAAGAACTATG	C524X synt R	T (15) TCAT GCTTTGATGACGCTTCTGTATC
V520F synt F	T (15) TTCA TCAAAGCAAGCAACTAGAAGA	V520F synt R	T (15) AGCT TCTGTATCTATATTCATCATAG
1609delCA synt F	T (15) TGTT TTCTCGGATTATGCTTGGCAC	1609delCA synt R	T (15) CAGA ACAGAAATGAATTTCTTCCACTG
1717 – 8G>A synt F	T (15) AGTA ATAGGACATCTCCAAGTTTGCA	1717 – 8G>A synt R	T (15) TAAA ATAGAAAATTAGAGAGTCACT
1784delG synt F	T (15) AGT CAACGAGCAAGAAATTTCTTTAGC	1784delG synt R	T (15) ACT CCACTCAGTGTGATTCACCTTTC
A559T synt F	T (15) ACA AGGTGAATAACTAATTTATTTGGT	A559T synt R	T (15) TTAA AGAAATCTTGTGCTGTTGACCT
Q552X synt F	T (15) TAA CGAGCAAGAAATTTCTTTAGCAAG	Q552X synt R	T (15) AACT TCCACTCAGTGTGATTTCCACCT
S549R(A>C) synt F	T (15) CGT GAGGTCAACGAGCAAGAAATTT	S549R(A>C) synt R	T (15) GCAG TGTGATTCTACCTTCTCCAAGA
S549R(T>G) synt F	T (15) GGG AGGTCAACGAGCAAGTATTTTC	S549R(T>G) synt R	T (15) CCT CAGTGTGATTTCCACCTTCTCCAA
L558S synt F	T (15) CAG CAAGGTGAATAACTAATTTATTTGG	L558S synt R	T (15) GAGA AAATCTTCGCTCGTTGACCTCC
1811 + 1.6 kb A>G synt F	T (15) TGTA GTAAAGTTACTATCAATCACAC	1811 + 1.6 kb A>G synt R	T (15) CATC TCAAGTACATGAGGATTTCTGT
1812 – 1G>A synt F	T (15) AAG CAGTATACAAAGATCTGATTG	1812 – 1G>A synt R	T (15) TAAA AAAGAAAAATGGAATTTAAATTA
D572N synt F	T (15) AACT CTCCTTTTGGATACCTAGATGT	D572N synt R	T (15) TTA ATAAATACAAATCAGCATCTTTG
P574H synt F	T (15) ATT TTGGATACCTAGATGTTTAAACA	P574H synt R	T (15) TGAG AGTCTAATAAATACAAATCAGC
1833delT synt F	T (15) ATT GTATTTATTAGACTCTCCTTTTG	1833delT synt R	T (15) CAAT CAGCATCTTTGTATACTGCTCT

Table 4. Continued

Primer name	Sense strand 5' → 3'	Name	Antisense strand 5' → 3'
Y563D synt F	T (15) G ACAAAGATGCTGATTGTATTTATT	Y563D synt R	T (15) C TACTGCTCTAAAAAGAAAATGGAAA
T582R synt F	T (15) G AGAAAAAGAAATATTTGAAAGGTAT	T582R synt R	T (15) C TTAAACATCTAGGTATCCAAAAGG
E585X synt F	T (15) T AAATATTTGAAAGGTATGTTCTTTG	E585X synt R	T (15) A TTTTTCTGTTAAAAACATCTAGGTAT
1898 + 5G>T synt F	T (15) T TTCTTTGAATACCTTACTTATAATTG	1898 + 5G>T synt R	T (15) A ATACCTTTCAAATATTTCTTTTCT
1924del7 synt F	T (15) C AGGATTTTGGTCACTTCTAAAAATGG	1924del7 synt R	T (15) C TGTTAGCCATCAGTTTACAGACACA
2055del9>A synt F	T (15) A CATGGGATGTGATTCCTTCGACCAA	2055del9>A synt R	T (15) T CTAAAGTCTGGCTGTAGATTTTGGA
D648V synt F	T (15) T TTCTTTTCGACCAATTTAGTGCAGAA	D648V synt R	T (15) A CACATCCCATGAGTTTGTAGCTAAA
K710X synt F	T (15) T AATTTTCGACCTTGTGCAAAAGACTCC	K710X synt R	T (15) A TCGTATAGAGTTGATTGGATTGAGA
I618T synt F	T (15) C TTTGATGAAGGTAGCAGCTATTTT	I618T synt R	T (15) G TTAATATTTTGTCACTTTCTTTTAA
R764X synt F	T (15) T GAAAGGAGCAGTCTGTCTTGAACCT	R764X synt R	T (15) A TGCCTGAAGCGTGGGCCAGTGCTG
Q685X synt F	T (15) T AATCTTTTAAACAGACTGGAGAGTT	Q685X synt R	T (15) A TTTTTTTGTCTGTCCAGGAGACA
R709X synt F	T (15) T GAAAAATTTCCATTGTGCAAAAGAC	R709X synt R	T (15) A TATAGAGTTGATTGGATTGAGAAAT
V754M synt F	T (15) A TGATCAGCACTGGCCCCAGCTTCA	V754M synt R	T (15) T CGTGATGCGAGGAGTATCGCCTCT
1949del84 synt F	T (15) A AAAACTACAGCCAGACTTTATCTC	1949del84 synt R	T (15) T TTTGTAGAGTGACCAAAATCCTAGT
2108delA synt F	T (15) G AATTCATCTCACTGAGACCTTAC	2108delA synt R	T (15) A TTCTCTCTTCTGCACATAAATGGTC
2176insC synt F	T (15) C CAAAAAACAATCTTTTAAACAGACTGGAGAG	2176insC synt R	T (15) G GTTCCTGTCCAGGAGACAGGAGCAT
2184delA synt F	T (15) C AAAAAACACATCTTTTAAACAGACTGG	2184delA synt R	T (15) G TTTTTTTGTCTGTCCAGGAGACAG
2105–2117 del13 synt F	T (15) A AACTGAGACCTTACACCGTTTCTCA	2105–2117 del13 synt R	T (15) T TTCTTCTGCACATAAATGGTCGAA
2307insA synt F	T (15) A AGAGGATTCGTATGAGCCTTTAGA	2307insA synt R	T (15) T TCGATGCCATTCATTTGTAAAGGA
W846X synt F	T (15) A AACACATACCTTCGATATATTACTGTCCAC	W846X synt R	T (15) T CATGTAGTCACTGCTGGTATGCTCT
2734G/AT synt F	T (15) T TAATTTTCTTCTGGCAGAGGTAAAGAT	2734G/AT synt R	T (15) T TAAGCACCAAAATAGCACAAATAT
2766del8 synt F	T (15) G GTGGCTCCTTGGAAAGTGAGTATTC	2766del8 synt R	T (15) C ACCAAAGAAGCAGCCACCTGGAATGG
2790 – 2A>G synt F	T (15) G GCACCTCCTCTTCAAGACAAGGGGAA	2790 – 2A>G synt R	T (15) C GTAAAGCAAAATAGGAATCGTTAAT
2991del32 synt F	T (15) T TCACACGTCGAAAGCAGGTACTTTG	2991del32 synt R	T (15) A AACATTTTGTGGTGTAAAATTTTCG
Q890X synt F	T (15) T AAACAAGAGGGAATAGTACTCATAG	Q890X synt R	T (15) A AGAGGAGTGCTGTGAAGCAAAATG
2869insG synt F	T (15) G ATTATGTGTTTTACATTTACGTGGG	2869insG synt R	T (15) C ACGAACCTGGTGTGGTGATAATCAC
3120G>A synt F	T (15) A GTATGTAAAAATAAGTACCGTTAAG	3120G>A synt R	T (15) T TGGATGAAGTCAAATATGGTAAGAG
3121 – 2A>T synt F	T (15) T GTTGTTTATTAATTTGTGATTGGAGCT	3121 – 2A>T synt R	T (15) A GTAAAGTCAAAAGAAACATGTTGGT
3132delTG synt F	T (15) T TGATTGCTTTTATATGTTGAGAGCATAT	3132delTG synt R	T (15) A ATTAAATAACAACATAGATCAAAG
3271delGG synt F	T (15) A TATGACAGTGAATGTGCGATACTCA	3271delGG synt R	T (15) A TTCAGATTCCAGTTGTTTGTAGTTGC
3171delC synt F	T (15) A CCTAGACTCTTTGTGCAACAGTGCC	3171delC synt R	T (15) A GGTTGTAAAACCTGCGACAACCTGCTA
3171insC synt F	T (15) C CCCACATCTTTGTTGCTACAGTGC	3171insC synt R	T (15) G GGGTTGTAAAACCTGCGACAACCTGCT
3199del6 synt F	T (15) G AGTGGCTTTTATATGTTGAGAGCATAT	3199del6 synt R	T (15) C CACCTGGCACTGTTGCAACCAAGATG
M1101K synt F	T (15) A GAGAATAGAAATGATTTTGTGCATC	M1101K synt R	T (15) T TTTGAACACGCGCAGTGTGACAG
G1061R synt F	T (15) C GACTATGACACTTCGTGCTTTCGG	G1061R synt R	T (15) G TTTTAAGCTTGTAAACAGATGAGTG
R1066L synt F	T (15) T TGCTTCGCGACGCGAGCCTTACTTT	R1066L synt R	T (15) A GAAGTGTCATAGTCCTTTTAAAGCT
R1070P synt F	T (15) C GCAGCCTTACTTTGAACTCTGTTG	R1070P synt R	T (15) G GTCCGAAGGCACGAGTGTCCATAG
L1077P synt F	T (15) C GTTCCACAAAGCTCTGAATTTACAT	L1077P synt R	T (15) G GAGTTTCAAAAGTAAGGCTGCCGTCC
W1089X synt F	T (15) A GTCTTGTACCTGTCAACACTGCGC	W1089X synt R	T (15) T AGTTGGCAGTATGTAAATTCAGAGC
L1093P synt F	T (15) C GTCAACACTCGCGTGTCCCAAATG	L1093P synt R	T (15) G GGTACAAGAACAGTTGGCAGTATG
W1098R synt F	T (15) C GGTTCCAAATGAGAATAGAAATGAT	W1098R synt R	T (15) G GGCGAGTTGTGACAGGTACAAGAAC
Q1100P synt F	T (15) C AATGAGAAATAGAAATGATTTTGTG	Q1100P synt R	T (15) G GGAAACGCGCAGTGTGTGACAGGT
D1152H synt F	T (15) C ATGTGGATAGCTTGGTAACTTTAT	D1152H synt R	T (15) G TATGCTGGAGTTTACAGCCCACTGC
R1158X synt F	T (15) T GATCTGTGAGCCGAGTCTTTAAGTT	R1158X synt R	T (15) A CATCTGAAATAAAAAATTAACACATT
S1196X synt F	T (15) G ACACGTGAAGAAAGATGACATCTGG	S1196X synt R	T (15) C AATCTCAATAATCATAACTTTTCA
3732delA synt F	T (15) G GAGATGACATCTGGCCCTCAGGGGG	3732delA synt R	T (15) C TCTCTCAGCTGTGAAATCTCTCAATAA
3791delC synt F	T (15) A GAAGGTGGAATGCCATATTAGAG	3791delC synt R	T (15) T TGTATTTTGTCTGTGAGATCTTTGAC
3821delT synt F	T (15) A TTCCCTCTCAATAAGTCCCTGCCAG	3821delT synt R	T (15) G AATGTTCTCTAATATGGCATTTCCA
Q1238X synt F	T (15) T AGAGGGTGAGATTGGAACACTGCTT	Q1238X synt R	T (15) A GCCAGGACTTATTGAGAAGGAAATG
S1255X (ex19)synt F	T (15) G TCTGGCCCTCAGGGGGCCAAATGAC	S1255X (ex19) synt R	T (15) C GTCACTCTTCTTCTCACTGTGAATTC
S1255X;L synt F	T (15) A AGCTTTTTTGTGAGACTACTGAACACT	S1255X;L synt R	T (15) T ATAACAAAGTAATCTTCCCTGATCC
3849 + 4A>G synt F	T (15) G ATTTTGAACTACTGCTTGTCTTTGTTA	3849 + 4A>G synt R	T (15) C CACCTCTGCGCCAGGACTTATTGAG
3850 – 1G>A synt F	T (15) A GTGGGCCCTCTTGGGAAGAAGCTGGAT	3850 – 1G>A synt R	T (15) T TATAAGGTAAAAGTGATGGGATCAC
3905insT synt F	T (15) T TTTTTTTGTGACTACTGAACACTGAA	3905insT synt R	T (15) A AAAAAGCTGATACAAAGTACTCT
3876delA synt F	T (15) C GGGAAGAGTACTTTTGTATCAGCTT	3876delA synt R	T (15) C GATCCAGTTCTTCCCAAGAGGCCA
G1244V synt F	T (15) T AGAAGTGGATCAGGGAAGAGTACT	G1244V synt R	T (15) A CCAAAGAGGCCACCTATAAGGTAAA
G1249E synt F	T (15) A GAAGAGTACTTTGTTATCAGCTTTT	G1249E synt R	T (15) T CTGATCCAGTTCTTCCCAAGAGGCC
S1251N synt F	T (15) A TACTTGTATTACGCTTTTTTGTGACTACTG	S1251N synt R	T (15) T CTTCCCTGTATGCAAGTTCTTCCAA
S1252P synt F	T (15) C CTTTGTTATCAGCTTTTTTGTGACT	S1252P synt R	T (15) G ACTCTTCCCTGATCCAGTTCTTCCC
D1270N synt F	T (15) A ATGGTGTGCTTGGGATTCATAAC	D1270N synt R	T (15) T GATCTGGATTTCTCTCTCAGTGTTC
W1282R synt F	T (15) C GGAGGAAAGCCTTTTGAGTGATACC	W1282R synt R	T (15) G CTGTTGCAAGTTTATTTGAATCCCCA
R1283K synt F	T (15) A GAAAGCCTTTGGAGTGATACCACAG	R1283K synt R	T (15) T TCCACTGTGCAAGTTTATTTGAATC
4005 + 1G>A synt F	T (15) A TAGGACAAAGGACTTAGCCAGAAAA	4005 + 1G>A synt R	T (15) T CTGTGATCACTCCAAAGGCTTTTC
4010del4 synt F	T (15) G TATTTTTTCTGGAACATTTAGAAAAAAGTGG	4010del4 synt R	T (15) A AAATACTTTCTATAGCAAAAAAGAAAAGAA
4016insT synt F	T (15) T TTTTTTCTGGAACATTTAGAAAAAAGTGG	4016insT synt R	T (15) A AAAAATAAATACTTTCTATAGCAAAAAAGAAAAGAA
CFTRdele21 synt F	T (15) T AGTAAGCTGTACTGAAATGAT	CFTRdele21 synt R	T (15) C CTATAGCAAAAAAGAAAAGAAAAGATATG
4382delA synt F	T (15) G AGAGACAAAGTGGCGAGTACGAT	4382delA synt R	T (15) C TCTATGACCTATGGAAATGGCTGTT

Bold, mutation allele of interest; bold and italicized, modified nucleotide.

Discussion

The CF APEX array system presents a new technology for comprehensive screening for CFTR mutations, one that offers several advantages. It is an integrated system with DNA microarray chip, multiplex primer extension on the array, and fluorescence imaging and data analysis.^{4,14} Because of the unprecedented number and selection of mutations on the CFTR array, this CF test has a higher mutation detection capability than assays currently available on-site in diagnostic laboratories. It also more than doubles the number of offered mutations on currently available panels at reference laboratories. The inclusion of mutations that have a higher frequency in specific ethnic populations increases this mutation detection capability for patients affected with cystic fibrosis (who have two mutations), as well as unaffected mutation carriers (who have one mutation). The CF APEX test is specific, reproducible, and technically robust. Test costs (including labor, supplies, and capital equipment) will ultimately depend on the scale of production and deployment, but the costs are clearly within the range of current commercial molecular diagnostic tests. In addition, the APEX system in general is applicable to many other molecular diagnostic tests for genetic and infectious diseases. For example, it has recently been developed for detection of more than 400 mutations in the ABCR (ABCA4) gene associated with five distinct retinal phenotypes¹⁵ and for detection of β -thalassemia and G6PD mutations.¹⁶ Thus, it is a "platform" technology that in turn amortizes labor and capital costs.

One concern with any platform that allows screening for a large number of mutations, of which some are rare in the general population, is the approach to quality control. Ideally, a positive control is run for each mutation every time the assay is performed. However, even with the widely used commercial CF screening panels that enable testing for approximately 25 mutations, this approach is prohibitively expensive and impractical. Thus, most laboratories have resorted to the inclusion of only a fraction of the positive controls in each assay so that all controls are assessed on a rotating basis. With the CF APEX array, the inclusion of positive controls is even more challenging because of the large number of mutations on the array and the fact that the majority of these are not commercially available. To perform the assay in a clinical diagnostic setting, native and/or synthetic positive control samples could be pooled for an assessment of the entire mutation list in only a few runs. This, of course, will require optimization to achieve ideal conditions. False-positive samples are avoided through several checkpoints in the assay: 1) a negative sample void of DNA is included in each run; 2) every sample is automatically assessed in duplicate; and 3) every sample, when technically possible, is evaluated in both the forward and reverse direction for direct result confirmation. Because it may remain impractical to obtain genomic DNA for routine quality control of all mutations tested, we expect that the use of synthetic samples may remain necessary. Synthetics are dependent on the correctness of the original sequence reported in the literature and cannot reflect the perfor-

mance of true test conditions, which include the possibility of genomic polymorphisms at or near the PCR primer sites and differences in PCR efficiency compared with genomic DNA. Thus, an initial full validation with genomic samples will be optimal.

Although the CF APEX array, including the 204 mutations presented herein, improves the mutation detection possibilities of first-line CFTR mutation screening, it is clear that many, if not most, mutations that contribute to CF in Hispanics, African Americans, and Asians remain unknown. It is essential that future research endeavors determine the prevalent mutations in each population, followed by inclusion of newly identified mutations on the CF APEX chip (or other diagnostic tests), so that appropriate genetic counseling and understanding of the pathogenesis and outcomes of CF can be achieved in all ethnic groups. Considering that more than 35 million Hispanic Americans reside in the United States and that this is the fastest growing segment of the population (<http://www.census.gov/>; <http://www.dhs.ca.gov/>), as well as one with a relatively high carrier frequency, it is imperative that more comprehensive screening routinely be provided. In addition, delayed diagnosis in children with CF can result in more severe health problems and can adversely affect prognosis. Diagnostic delay is more frequently observed in ethnic minorities.¹⁷

In addition to CF carrier screening, demand for molecular CFTR gene testing has been growing for another reason as well. It is increasingly recognized that mutations in the CFTR gene play a role in a variety of diseases other than classic CF, including CBAVD,¹⁸ chronic pancreatitis,¹⁹ disseminated bronchiectasis,²⁰ and allergic pulmonary aspergillosis.²¹ The CF APEX array can serve as the initial research tool to screen for mutations in these diseases, coupled with other more comprehensive but labor- and capital-intensive technologies if results are negative. As other important disease-contributive CFTR mutations are identified, these can be added to the CF APEX array for routine analysis.

Acknowledgments

We thank Dr. Tiina Kahre (Molecular Diagnostics Centre of United Laboratories, Tartu University Clinics) and Dr. Milan Macek, Jr. (Charles University, Institute of Biology and Medical Genetics, Department of Molecular Genetics, CF Center, Prague, Czech Republic) for the generous contribution of DNA samples.

References

1. Grody WW, Cutting GR, Klinger KW, Richards CS, Watson MS, Desnick RJ; Subcommittee on Cystic Fibrosis Screening, Accreditation of Genetic Services Committee, ACMG. American College of Medical Genetics: Laboratory standards and guidelines for population-based cystic fibrosis carrier screening. *Genet Med* 2001; 3:149-154
2. Watson MS, Cutting GR, Desnick RJ, Driscoll DA, Klinger K, Mennuti M, Palomaki GE, Popovich BW, Pratt VM, Rohlfis EM, Strom CM,

- Richards CS, Witt DR, Grody WW: Cystic fibrosis population carrier screening: 2004 revision of American College of Medical Genetics mutation panel. *Genet Med* 2004, 6:387–391
3. Girondon-Boulandet E, Cazeneuve C, Goossens M: Screening practices for mutations in the CFTR gene ABCC7. *Hum Mutat* 2000, 15:135–149
4. Kurg A, Tonisson N, Georgiou I, Shumaker J, Tollett J, Metspalu A: Arrayed primer extension: solid-phase four-color DNA resequencing and mutation detection technology. *Genet Test* 2000, 4:1–7
5. Shumaker JM, Metspalu A, Caskey CT: Mutation detection by solid phase primer extension. *Hum Mutat* 1996, 7:346–354
6. Highsmith WE, Burch LH, Zhou Z, Olsen JC, Boat TE, Spock A, Gorvoy JD, Quittel L, Friedman KJ, Silverman LM, et al. A novel mutation in the cystic fibrosis gene in patients with pulmonary disease but normal sweat chloride concentrations. *N Engl J Med* 1994, 331:974–980
7. Wong LJ, Wang J, Zhang YH, Hsu E, Heim RA, Bowman CM, Woo MS: Improved detection of CFTR mutations in Southern California Hispanic CF patients. *Hum Mutat* 2001 18:296–307
8. Mercier B, Raguenes O, Estivill X, Morral N, Kaplan GC, McClure M, Grebe TA, Kessler D, Pignatti PF, Marigo C, et al. Complete detection of mutations in cystic fibrosis patients of Native American origin. *Hum Genet* 1994 94:629–632
9. Zielenski J, Rozmahel R, Bozon D, Kerem B, Grzelczak Z, Riordan JR, Rommens J, Tsui LC: Genomic DNA sequence of the cystic fibrosis transmembrane conductance regulator (CFTR) gene. *Genomics* 1991, 10:214–228
10. Wong LJ, Wang J, Woo M, Hsu E, Bowman CM: A novel mutation detected by temporal temperature gradient gel electrophoresis led to the confirmative prenatal diagnosis of a Hispanic CF family. *Prenat Diagn* 2000, 20:807–810
11. Wang J, Bowman MC, Hsu E, Wertz K, Wong LJ: A novel mutation in the CFTR gene correlates with severe clinical phenotype in seven Hispanic patients. *J Med Genet* 2000, 37:215–218
12. Wang J, Bowman CM, Wong LJ: A novel CFTR frame-shift mutation, 935delA, in two Hispanic cystic fibrosis patients. *Mol Genet Metab* 2000, 70:316–321
13. Orozco L, Velazquez R, Zielenski J, Tsui LC, Chavez M, Lezana JL, Saldana Y, Hernandez E, Carnevale A: Spectrum of CFTR mutations in Mexican cystic fibrosis patients: identification of five novel mutations (W1098C, 846delT, P750L, 4160insGGGG and 297–1G→A). *Hum Genet* 2000, 106:360–365
14. Tonisson N, Kurg A, Kaasik K, Lohmussaar E, Metspalu A: Unravelling genetic data by arrayed primer extension. *Clin Chem Lab Med* 2000 38:165–170
15. Jaakson K, Zernant J, Kulm M, Hutchinson A, Tonisson N, Glavac D, Ravnik-Glavac M, Hawlina M, Meltzer MR, Caruso RC, Testa F, Maugeri A, Hoyng CB, Gouras P, Simonelli F, Lewis RA, Lupski JR, Cremers FPM, Allikmets R: Genotyping microarray (Gene Chip) for the ABCR (ABCA4) gene. *Hum Mutat* 2003, 22:395–403
16. Gemignani F, Perra C, Landi S, Canzian F, Kurg A, Tonisson N, Galanello R, Cao A, Metspalu A, Romeo G: Reliable detection of b-Thalassemia and G6PD mutations by a DNA microarray. *Clin Chem* 2002, 48:2051–2054
17. Spencer DA, Venkataraman M, Weller PH: Delayed diagnosis of cystic fibrosis in children from ethnic minorities. *Lancet* 1993, 342:238
18. Chillon M, Casals T, Mercier B, Bassas L, Lissens W, Silber S, Romey MC, Ruiz-Romero J, Verlingue C, Claustres M, Nunes V, Ferec C, Estivill X: Mutations in the cystic fibrosis gene in patients with congenital absence of the vas deferens. *N Engl J Med* 1995, 332:1475–1480
19. Pezzilli R, Morselli-Labate AM, Mantovani V, Romboli E, Selva P, Migliori M, Corinaldesi R, Gullo L: Mutations of the CFTR gene in pancreatic disease. *Pancreas* 2003, 4:332–336
20. Pignatti PF, Bombieri C, Marigo C, Benetazzo M, Luisetti M: Increased incidence of cystic fibrosis gene mutations in adults with disseminated bronchiectasis. *Hum Mol Genet* 1995, 4:635–639
21. Eaton TE, Weiner Miller P, Garrett JE, Cutting GR: Cystic fibrosis transmembrane conductance regulator gene mutations: do they play a role in the aetiology of allergic bronchopulmonary aspergillosis? *Clin Exp Allergy* 2002, 32:756–761